

WHAT IS CLAIMED IS:

1. A morphological analyzer comprising:

a hypothesis generator for applying a prescribed method of morphological analysis to a text and generating one or more hypotheses as candidate results of the morphological analysis, each hypothesis being a word string with part-of-speech tags, the part-of-speech tags including form information for parts of speech having forms;

a model storage facility storing information for a plurality of part-of-speech n-gram models, at least one of the part-of-speech n-gram models including information about the forms of the parts of speech;

a probability calculator for finding a probability that each said hypothesis will appear in a large corpus of text by using a weighted combination of the information for the part-of-speech n-gram models stored in the model storage facility; and

a solution finder for finding a solution among said hypotheses, based on the probabilities generated by the probability calculator.

2. The morphological analyzer of claim 1, wherein said at least one of the part-of-speech n-gram models including information about forms of parts of speech is a hierarchical part-of-speech n-gram model.

3. The morphological analyzer of claim 2, wherein the hierarchical part-of-speech n-gram model calculates a product of a conditional probability $P(w_i | t_i)$ of occurrence of a word w_i given its part of speech t_i , a conditional probability $P(t_i^{\text{form}} | t_i^{\text{pos}})$ of occurrence of the part of speech t_i^{pos} of said word w_i in a form t_i^{form} shown by said word w_i , and a conditional probability $P(t_i^{\text{pos}} | t_{i-N+1} \dots t_{i-1})$ of

occurrence of the part of speech t_i^{pos} of said word w_i following a part-of-speech tag string $t_{i-N+1} \dots t_{i-1}$ indicating parts of speech of $N - 1$ preceding words, where N is a positive integer.

4. The morphological analyzer of claim 1, wherein at least one of the part-of-speech n -gram models is a lexicalized part-of-speech n -gram model.

5. The morphological analyzer of claim 4, wherein the lexicalized part-of-speech n -gram model calculates a product of a conditional probability $P(w_i | t_i)$ of occurrence of a word w_i given its part of speech t_i and a conditional probability $P(t_i | w_{i-N+1} t_{i-N+1} \dots w_{i-1} t_{i-1})$ of occurrence of the part of speech t_i of said word w_i following $N - 1$ words $w_{i-N+1} \dots w_{i-1}$ having respective parts of speech $t_{i-N+1} \dots t_{i-1}$, where N is a positive integer.

6. The morphological analyzer of claim 4, wherein the lexicalized part-of-speech n -gram model calculates a conditional probability $P(w_i t_i | t_{i-N+1} \dots t_{i-1})$ of occurrence of a word w_i having a part of speech t_i following a string of $N - 1$ parts of speech $t_{i-N+1} \dots t_{i-1}$, where N is a positive integer.

7. The morphological analyzer of claim 4, wherein the lexicalized part-of-speech n -gram model calculates a conditional probability $P(w_i t_i | w_{i-N+1} t_{i-N+1} \dots w_{i-1} t_{i-1})$ of occurrence of a word w_i having a part of speech t_i following a string of $N - 1$ words $w_{i-N+1} \dots w_{i-1}$ having respective parts of speech $t_{i-N+1} \dots t_{i-1}$, where N is a positive integer.

8. The morphological analyzer of claim 1, wherein at least one of the part-of-speech n -gram models stored in the model

storage facility is a class part-of-speech n-gram model.

9. The morphological analyzer of claim 8, wherein the class part-of-speech n-gram model calculates a product of a conditional probability $P(w_i|t_i)$ of occurrence of a word w_i given its part of speech t_i and a conditional probability $P(t_i|c_{i-N+1}t_{i-N+1}...c_{i-1}t_{i-1})$ of occurrence of said part of speech t_i following a string of $N - 1$ words assigned to respective classes $c_{i-N+1}...c_{i-1}$ with respective parts of speech $t_{i-N+1}...t_{i-1}$, where N is a positive integer.

10. The morphological analyzer of claim 8, wherein the class part-of-speech n-gram model calculates a product of a conditional probability $P(w_i t_i | c_{i-N+1} t_{i-N+1} ... c_{i-1} t_{i-1})$ of occurrence of a word w_i having a part of speech t_i following a string of $N - 1$ words in respective classes $c_{i-N+1}...c_{i-1}$ with respective parts of speech $t_{i-N+1}...t_{i-1}$, where N is a positive integer.

11. The morphological analyzer of claim 8, wherein the class part-of-speech n-gram model is trained from both a part-of-speech tagged corpus and a part-of-speech untagged corpus.

12. The morphological analyzer of claim 1, further comprising a weight calculation unit using a leave-one-out method to calculate weights of the part-of-speech n-gram models.

13. A method of morphological analysis comprising:
 applying a prescribed method of morphological analysis to a text and generating one or more hypotheses as candidate results of the morphological analysis, each hypothesis being a word string with part-of-speech tags, the part-of-speech

tags including form information for parts of speech having forms;

calculating probabilities that each said hypothesis will appear in a large corpus of text by using a weighted combination of a plurality of part-of-speech n-gram models, at least one of the part-of-speech n-gram models including information about forms of parts of speech; and

finding a solution among said hypotheses, based on said probabilities.

14. The method of claim 13, wherein said at least one of the part-of-speech n-gram models including information about forms of parts of speech is a hierarchical part-of-speech n-gram model.

15. The method of claim 14, wherein the hierarchical part-of-speech n-gram model calculates a product of a conditional probability $P(w_i | t_i)$ of occurrence of a word w_i given its part of speech t_i , a conditional probability $P(t_i^{\text{form}} | t_i^{\text{pos}})$ of occurrence of the part of speech t_i^{pos} of said word w_i in a form t_i^{form} shown by said word w_i , and a conditional probability $P(t_i^{\text{pos}} | t_{i-N+1} \dots t_{i-1})$ of occurrence of the part of speech t_i^{pos} of said word w_i following a part-of-speech tag string $t_{i-N+1} \dots t_{i-1}$ indicating parts of speech of $N - 1$ preceding words, where N is a positive integer.

16. The method of claim 13, wherein at least one of the part-of-speech n-gram models is a lexicalized part-of-speech n-gram model.

17. The method of claim 13, wherein at least one of the part-of-speech n-gram models is a class part-of-speech n-gram model.

18. The method of claim 17, further comprising training the class part-of-speech n-gram model from both a part-of-speech tagged corpus and a part-of-speech untagged corpus.

19. The method of claim 13, further comprising using a leave-one-out method to calculate weights of the part-of-speech n-gram models.

20. A machine-readable medium storing a program comprising instructions that can be executed by a computing device to carry out morphological analysis by the method of claim 13.